

# Well-posedness of the permutation problem in sparse filter estimation with lp minimization

Alexis Benichoux, Prasad Sudhakar, Frédéric Bimbot, Rémi Gribonval

► **To cite this version:**

Alexis Benichoux, Prasad Sudhakar, Frédéric Bimbot, Rémi Gribonval. Well-posedness of the permutation problem in sparse filter estimation with lp minimization. [Research Report] RR-7782, INRIA. 2011, pp.20. hal-00640198

**HAL Id: hal-00640198**

**<https://hal.inria.fr/hal-00640198>**

Submitted on 10 Nov 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Well-posedness of the permutation problem in sparse filter estimation with $\ell^p$ minimization

Alexis Benichoux, Prasad Sudhakar,  
Frédéric Bimbot, and Rémi Gribonval

**RESEARCH  
REPORT**

**N° 7782**

November 2011

Project-Team METISS





## Well-posedness of the permutation problem in sparse filter estimation with $\ell^p$ minimization

Alexis Benichoux, Prasad Sudhakar,  
Frédéric Bimbot, and Rémi Gribonval

Project-Team METISS

Research Report n° 7782 — November 2011 — 20 pages

**Abstract:** Convolutional source separation is often done in two stages: 1) estimation of the mixing filters and 2) estimation of the sources. Traditional approaches suffer from the ambiguities of arbitrary permutations and scaling in each frequency bin of the estimated filters and/or the sources, and they are usually corrected by taking into account some special properties of the filters/sources. This paper focusses on the filter permutation problem in the absence of scaling, investigating the possible use of the temporal sparsity of the filters as a property enabling permutation correction. Theoretical and experimental results highlight the potential as well as the limits of sparsity as an hypothesis to obtain a well-posed permutation problem.

**Key-words:** sparse filter, convolutional blind source separation, permutation ambiguity,  $\ell^p$  minimization, Hall's Marriage Theorem, bi-stochastic matrix

**RESEARCH CENTRE  
RENNES – BRETAGNE ATLANTIQUE**

Campus universitaire de Beaulieu  
35042 Rennes Cedex

## Caractère bien posé du problème de permutation pour l'estimation des filtres parcimonieux par minimisation $\ell^p$

**Résumé :** La séparation de source des mélanges convolutifs se fait souvent en deux étapes : 1) estimation des filtres de mélange et 2) estimation des sources. Les approches classiques souffrent d'ambiguïtés de permutation et de facteur d'échelle arbitraire pour chaque fréquence des filtres et/ou des sources estimés. Ces ambiguïtés sont habituellement corrigées en prenant en compte des propriétés particulières des filtres/sources. Cet article se concentre sur le problème de permutation des filtres en l'absence de facteur d'échelle, en explorant l'utilisation potentielle de la parcimonie temporelle des filtres pour résoudre le problème de permutation. Les résultats théoriques et expérimentaux soulignent tant le potentiel que les limites de l'hypothèse de parcimonie pour obtenir un problème bien posé.

**Mots-clés :** Filtres parcimonieux, séparation aveugle de sources, mélange convolutif, ambiguïté de permutation, théorème de mariage de Hall, matrice bi-stochastique

## 1 Introduction

Blind source separation and blind source localization are ubiquitous problems in signal processing, with applications ranging from wireless telecommunications to underwater acoustics and sound enhancement.

These problems can be considered as reasonably well understood and solved in simple linear instantaneous settings, where tools such as Independent Component Analysis, as well as techniques exploiting source sparsity, are now mature. However, the convolutive source localization / separation problem remains much more challenging. In particular, without further assumption than statistical independence between sources, the problem is known to be ill-posed because of the so-called frequency permutation (and scaling) problem: at best, one can hope to estimate for each frequency (up to a source and frequency dependent scaling factor) the collection of frequency components of all sources (and of the associated mixing filters); but one cannot match the estimated frequency components from different subbands to globally identify the sources (and mixing filters).

Several practical approaches have been proposed to solve the permutation and scaling problems in practice, by exploiting various properties of either the mixing filters or the sources to match different frequency subbands. While some of these methods may succeed in practice for certain types of sources / filters, there is no known theory guaranteeing the well-posedness of the permutation and scaling problem under appropriate assumptions.

This paper contributes to fill this gap, by providing well-posedness guarantees for the permutation problem under *sparsity* assumptions on the mixing filters. Sparse filters, associated to impulse responses corresponding to a limited set of echoes, are typically encountered in a number of underwater communication channels [1] or wireless telecommunications scenarios [2, 3] which are relevant for blind source localization and separation, and the theoretical results achieved in this paper indicate that this property can potentially be exploited for blind estimation in this context.

### 1.1 Problem formulation and notations

Let  $x_i[t]$ ,  $1 \leq i \leq M$  be  $M$  mixtures of  $N$  source signals  $s_j[t]$ , resulting from the convolution with filters  $a_{ij}[t]$  of length  $L$  such that:

$$x_i[t] = \sum_{j=1}^N (a_{ij} \star s_j)[t], \quad 1 \leq i \leq M, \quad (1)$$

where  $\star$  denotes convolution. The filter  $a_{ij}[t]$  typically models the impulse response between the  $j^{\text{th}}$  source and the  $i^{\text{th}}$  sensor. By abuse of notation,  $\mathbf{F}a_{ij} = \{a_{ij}[\omega]\}_{0 \leq \omega < L}$  denotes the discrete Fourier transform of the filter seen as a vector  $a_{ij} = \{a_{ij}[t]\}_{0 \leq t < L} \in \mathbb{C}^L$ . Also, the mixing equation (1) can be rewritten as  $\mathbf{X} = \mathbf{A} \star \mathbf{S}$ , with  $\mathbf{A}$  the matrix of filters

$$\mathbf{A} := (\{a_{ij}[t]\}_{0 \leq t < L})_{1 \leq i \leq M, 1 \leq j \leq N},$$

$\mathbf{X}$  the observation matrix and  $\mathbf{S}$  the source matrix.

In this context, blind filter estimation refers to the problem of obtaining estimates of the filters  $\mathbf{A}$  from the mixtures  $\mathbf{X}$ , without any explicit knowledge

about the sources  $\mathbf{S}$ . Mixing filters estimation is relevant for several purposes such as deconvolution, source localization, etc. [4]. It also has a relationship with the problem of Multiple-Input-Multiple-Output (MIMO) system identification in communications engineering [5].

## 1.2 Frequency domain filter estimation

Estimating the mixing parameters is made easier when all filters are instantaneous, that is to say of length  $L = 1$ , as the convolution product in (1) is replaced by the usual product. However, things get complicated in the general setting of convolutive mixtures.

A common approach for filter estimation then relies on the transformation of the mixing model in Eq. (1) into the time-frequency domain, converting a single convolutive filter estimation problem into several complex instantaneous filter estimation problems. Using standard techniques for instantaneous mixing parameter estimation [6], complex mixing filter coefficients

$$\tilde{\mathbf{A}}[\omega] = \{\tilde{a}_{ij}[\omega]\}_{1 \leq i \leq M, 1 \leq j \leq N}$$

are estimated for each frequency bin  $0 \leq \omega < L$ .

## 1.3 Permutation and scaling ambiguities

Without further assumption on either the filters  $a_{ij}[t]$  or the sources  $s_j[t]$ , one can at best hope to find an estimation  $\hat{\mathbf{A}} = (\hat{a}_{ij})$  where for every frequency  $\omega$  we have

$$\tilde{a}_{ij}[\omega] = \lambda_j[\omega] a_{i\sigma_\omega(j)}[\omega], \quad (2)$$

with  $\lambda_j[\omega]$  a scaling ambiguity and  $\sigma_\omega \in \mathfrak{S}_N$  a permutation ambiguity, where  $\mathfrak{S}_N$  is the set of permutations of the integers between one and  $N$ . Several methods [7] attempt to solve for these ambiguities by exploiting properties of either the sources  $\mathbf{S}$  or the filters  $\mathbf{A}$  [8, 9, 10].

## 1.4 Exploiting sparsity to solve the permutation ambiguity

The focus of this article is the use of the *sparsity* of  $\mathbf{A}$  in the time domain to find  $\sigma_0 \dots \sigma_{L-1} \in \mathfrak{S}_N$ , assuming the scaling is solved, i.e.,  $\lambda_j[\omega] = 1$ .

Assuming that  $\mathbf{A}$  is sparse means that each filter  $a_{ij}$  has few nonzero coefficients, as measured by the  $\ell^0$  pseudo-norm

$$\|a_{ij}\|_0 := \#\{0 \leq t < L, a_{ij}[t] \neq 0\} = \sum_t |a_{ij}[t]|^0.$$

The approach considered in this article is to seek permutations  $\hat{\sigma}_0, \dots, \hat{\sigma}_{L-1}$  yielding the sparsest estimated time-domain matrix of filters  $\hat{\mathbf{A}} = (\hat{a}_{ij})$  where  $\hat{a}_{ij}[\omega] := \tilde{a}_{i\hat{\sigma}_\omega(j)}[\omega]$ . Besides the  $\ell^0$  pseudo-norm  $\|\hat{\mathbf{A}}\|_0 := \sum_{ij} \|\hat{a}_{ij}\|_0$ , the following  $\ell^p$  quasi-norms will be used to quantify the sparsity of  $\hat{\mathbf{A}}$ :

$$\|\hat{\mathbf{A}}\|_p^p := \sum_{ij} \|\hat{a}_{ij}\|_p^p = \sum_{ijt} |\hat{a}_{ij}[t]|^p, \quad 0 < p \leq 1.$$

## 1.5 Main results

Our main result (Theorem 2) is a theoretical guarantee that when the filter length  $L$  is prime,  $k$ -sparse filters (i.e., such that  $\|a_{ij}\|_0 \leq k$ ) uniquely minimize the  $\ell^0$  norm of  $\mathbf{A}$  (up to a global permutation) if  $\frac{k}{L} \leq \alpha(N)$ , where  $N$  is the number of sources. To reach this bound we exploit uncertainty principles as well as the bistochastic structure of the problem through an apparently new quantitative result on bistochastic matrices (Lemma 2).

## 1.6 Structure of the paper

The main theorems are stated in Section 2, and the main ingredients of their proofs are described in Section 3. In Section 4 we discuss the strength of the assumptions used in the theorems, and how much these could be relaxed. In Sec. 5, a naive combinatorial  $\ell^p$  minimization algorithm is proposed to resolve filter permutations and used for Monte-Carlo simulations. We conclude with a discussion of the potential, as well as the limits, of sparsity as a hypothesis to solve permutation problems, in connection with the theoretical and empirical results. All proofs are gathered in the appendix.

## 2 Theoretical guarantees

Given an  $M \times N$  filter matrix  $\mathbf{A}$ , made of filters of length  $L$ , and an  $L$ -tuple  $(\sigma_0, \dots, \sigma_{L-1}) \in \mathfrak{S}_N$  of permutations, we let  $\tilde{\mathbf{A}}$  be the matrix obtained from  $\mathbf{A}$  by applying the permutations in the frequency domain, as in (2), without scaling ( $\lambda_j[\omega] = 1$ ).

The effect of the permutations is said to coincide with that of a global permutation  $\pi \in \mathfrak{S}_N$  of the columns of  $\mathbf{A}$  if  $\tilde{a}_{ij} = a_{i\pi(j)}$ ,  $\forall i, j$ , or equivalently in the frequency domain:

$$\tilde{a}_{ij}[\omega] := a_{i\sigma_\omega(j)}[\omega] = a_{i\pi(j)}[\omega], \quad 0 \leq \omega < L, \quad \forall i, j.$$

This is denoted  $\mathbf{A} \equiv \tilde{\mathbf{A}}$ . First, we show that for filters with disjoint time-domain supports, permutations cannot decrease the  $\ell^p$  norm,  $0 \leq p \leq 1$ :

**Theorem 1** *Let  $\Gamma_{ij} \subset \{0, \dots, L-1\}$  be the time-domain support of  $a_{ij}$ . Suppose that for all  $i$  and  $j_1 \neq j_2$  we have*

$$\Gamma_{i,j_1} \cap \Gamma_{i,j_2} = \emptyset. \quad (3)$$

*Then, for  $0 \leq p \leq 1$ , we have  $\|\tilde{\mathbf{A}}\|_p \geq \|\mathbf{A}\|_p$ .*

Note that filters with disjoint supports need not be very sparse:  $M$  filters of length  $L$  can have disjoint supports provided that  $\max_j \|a_{ij}\|_0 \leq L/M$ . Yet, disjointness of filter supports is a strong assumption, and Theorem 1 only indicates that frequency permutations cannot decrease the  $\ell^p$  norm. Thus, the minimum value of the  $\ell^p$  norm might not be uniquely achieved (up to a global permutation). In our main result, we consider  $k$ -sparse filters of prime length, and  $p = 0$ :



**Theorem 2** Let  $\mathbf{A}$  be an  $M \times N$  matrix of filters of prime length  $L$ . Assume that

$$\max_{ij} \|a_{ij}\|_0 \leq k, \quad (4)$$

where

$$\frac{k}{L} \leq \alpha(N) := \begin{cases} \frac{2}{N(N+2)} & \text{if } N \text{ is even,} \\ \frac{2}{(N+1)^2} & \text{if } N \text{ is odd.} \end{cases} \quad (5)$$

Then, up to a global permutation,  $\mathbf{A}$  uniquely minimises the  $\ell^0$  pseudo-norm among all possible frequency permutations.

### 3 Main elements of the proof of Theorem 2

The proof of Theorem 2 relies on a measure of the “amount” of incurred permutation, on uncertainty principles, and on combinatorial arguments related to bi-stochastic matrices, involving Hall’s Marriage Theorem.

#### 3.1 Measures of the amount of incurred permutations

To measure the “amount” of incurred permutation, one can count the number of frequency bands where a non-trivial permutation is incurred, with respect to the best matching reference global permutation  $\pi$ , i.e.,  $\min_{\pi} \#\{\omega, \sigma_{\omega} \neq \pi\}$ . However, this generally yields the maximum count  $L - 1$ .

An alternative is to count the “size” of the incurred permutations, given a reference global permutation  $\pi$ , as the maximum number of frequencies where each estimated filter actually differs from the (globally permuted) original filters, yielding:

$$\Delta(\tilde{\mathbf{A}}, \mathbf{A}|\pi) := \max_{i,j} \|\mathbf{F}(\tilde{a}_{ij} - a_{i\pi(j)})\|_0 \quad (6)$$

$$\Delta(\tilde{\mathbf{A}}, \mathbf{A}) := \min_{\pi \in \mathfrak{S}_N} \Delta(\tilde{\mathbf{A}}, \mathbf{A}|\pi). \quad (7)$$

Note that  $\Delta(\tilde{\mathbf{A}}, \mathbf{A}) = 0$  iff  $\tilde{\mathbf{A}} \equiv \mathbf{A}$ .

#### 3.2 Exploitation of an uncertainty principle

With this notation, we have the following Lemma:

**Lemma 1** Assume that  $\tilde{\mathbf{A}} \not\equiv \mathbf{A}$ , that  $L$  is a prime integer, and that (4) holds with

$$2k + \Delta \leq L. \quad (8)$$

Then  $\|\tilde{\mathbf{A}}\|_0 > \|\mathbf{A}\|_0$  and  $\|\tilde{a}_{ij}\|_0 \geq \|a_{ij}\|_0, \forall i, j$ . The latter inequality is strict when  $\tilde{a}_{ij} \neq a_{ij}$ . For a general  $L$  (not necessarily prime), the same conclusions hold when the assumption (8) is replaced with

$$2k \cdot \Delta < L. \quad (9)$$

The skilled reader will rightly sense the role of uncertainty principles [11, 12, 13, Theorem 1] in the above lemma.

### 3.3 Combinatorial arguments

Using Lemma 1 with prime  $L$ , a simple combinatorial argument can be used to obtain a weakened version of Theorem 2, with the more conservative constant  $\alpha'(N) := 1/2N!$ : by the pigeonhole principle, for any  $L$ -tuple of frequency permutations among  $N$  sources, at least  $L/N!$  permutations are identical; as a result,  $\Delta(\tilde{\mathbf{A}}, \mathbf{A})$  is universally bounded from above by  $L - L/N!$ ; hence if  $k \leq L/2N!$  we obtain  $2k + \Delta \leq L$  and we can conclude thanks to Lemma 1.

The proof of Theorem 2 with the constant  $\alpha(N)$  exploits a stronger universal upper bound  $\Delta(\tilde{\mathbf{A}}, \mathbf{A}) \leq L(1 - 2\alpha(N))$ , obtained through an apparently new quantitative application of Hall's Marriage Theorem [14] to bi-stochastic matrices.

**Definition 1 (Bi-stochastic matrix)** *An  $N \times N$  matrix  $\mathbf{B}$  is called bi-stochastic if all its entries are non-negative, and the sum of the entries over each row as well as the sum of the entries over each column is one.*

**Lemma 2** *Let  $\mathbf{B}$  be an  $N \times N$  bi-stochastic matrix: there exists a permutation matrix  $\mathbf{P}$  such that all the entries of  $\mathbf{B}$  on the support of  $\mathbf{P}$  exceed the threshold*

$$2\alpha(N) = \begin{cases} \frac{4}{N(N+2)} & \text{if } N \text{ is even,} \\ \frac{4}{(N+1)^2} & \text{if } N \text{ is odd.} \end{cases} \quad (10)$$

**Corollary 1** *Let  $\sigma_0, \dots, \sigma_{L-1} \in \mathfrak{S}_N$  be  $L$  permutations. There exists a global permutation  $\pi$  such that*

$$C_{j\pi(j)} = \#\{\ell : \sigma_\ell(j) = \pi(j)\} \geq 2L\alpha(N), \quad \forall 1 \leq j \leq N.$$

## 4 Discussion

The reader may have noticed that Theorem 2, while dropping the *disjoint support* assumption from Theorem 1, introduces new restrictions: the assumption that  $L$  is prime, and the restriction to  $p = 0$  compared to  $0 \leq p \leq 1$  in Theorem 1. How important are these restrictions? Could they be relaxed while exploiting sparsity together with the *disjoint support* assumption? This is discussed in this section.

### 4.1 Extending Theorem 2 to non-prime filter length $L$ ?

As indicated by Lemma 3 below, for even  $L \geq 4$ , there exists sparse matrices of filters that are the sparsest *but not unique* (even up to a global permutation) solution of the considered problem: certain frequency permutations provide an *equally sparse but not equivalent solution*.

**Lemma 3** *For any integer  $k$  such that  $2k$  divides  $L$ , there exists a matrix of  $k$ -sparse filters  $\mathbf{A}$  and a set of  $L/2k$  frequency permutations resulting in  $\tilde{\mathbf{A}} \neq \mathbf{A}$  such that for all  $0 \leq p < \infty$ :  $\|\tilde{\mathbf{A}}\|_p = \|\mathbf{A}\|_p$ , and*

$$\|\tilde{a}_{ij}\|_p = \|a_{ij}\|_p, \quad \forall i, j. \quad (11)$$

*We have  $2k \cdot \Delta(\tilde{\mathbf{A}}, \mathbf{A}) = L$ .*

The fact that the filter matrices  $\mathbf{A}$  and  $\tilde{\mathbf{A}}$  satisfy  $2k \cdot \Delta(\tilde{\mathbf{A}}, \mathbf{A}) = L$  shows the sharpness of Lemma 1 for the case when  $L$  is even: the strict inequality in (9) cannot be improved.

Specializing Lemma 3 to  $k = 1$  for even  $L \geq 4$  yields ideally 1-sparse filters  $a_{ij}$  and a set of  $L/2$  frequency permutations such that:  $\tilde{a}_{ij}$  are 1-sparse;  $\tilde{\mathbf{A}}$  is not equivalent to  $\mathbf{A}$  and cannot be discriminated from it by any  $\ell^p$  norm.

## 4.2 Stronger guarantees with disjoint supports *and* sparsity ?

Could one get improved results by combining the disjoint support assumptions from Theorem 1 and the sparsity assumption from Theorem 2 ? For even  $L \geq 4$ , Lemma 4 below indicates the existence of sparse matrices of filters *with disjoint supports* that are the sparsest but not unique (even up to a global permutation) solution of the considered problem: certain frequency permutations of “size”  $\Delta = L/2k$  provide an equally good but not equivalent solution.

**Lemma 4** *For any integers  $k' < k \leq L/2$  such that  $2k'$  divides  $L$ , there exist a matrix of  $k$ -sparse filters  $\mathbf{A}$  with disjoint supports (3), and a set of  $L/2k'$  frequency permutations resulting in  $\tilde{\mathbf{A}} \neq \mathbf{A}$ , such that for all  $0 \leq p \leq \infty$ :  $\|\tilde{\mathbf{A}}\|_p = \|\mathbf{A}\|_p$  and*

$$\|\tilde{a}_{ij}\|_p = \|a_{ij}\|_p, \quad \forall i, j. \quad (12)$$

We have  $2k' \cdot \Delta(\tilde{\mathbf{A}}, \mathbf{A}) = L$ .

Specializing Lemma 4 to  $k' = 1$  and  $k = 2$  for even  $L \geq 4$  yields 2-sparse filters  $a_{ij}$  and a set of  $L/2$  frequency permutations such that:  $\tilde{a}_{ij}$  are 2-sparse;  $\tilde{\mathbf{A}}$  is not equivalent to  $\mathbf{A}$  and cannot be discriminated from it by any  $\ell^p$  norm.

This shows that even by adding the disjoint support assumption, for even  $L \geq 4$ , there is little margin to improve Lemma 1: at best, one can hope to replace the strict inequality in (9) with a large one. Can this actually be done ? This is partially answered by the following results:

**Lemma 5** *Assume that  $\tilde{\mathbf{A}} \neq \mathbf{A}$ , that (4) holds with*

$$2k \cdot \Delta(\tilde{\mathbf{A}}, \mathbf{A}) = L \quad (13)$$

*and that the filters in  $\mathbf{A}$  have disjoint supports (3). Then, either  $\|\tilde{\mathbf{A}}\|_0 > \|\mathbf{A}\|_0$ , or each row of  $\tilde{\mathbf{A}}$  is obtained by permuting pairs of distinct filters  $a_{ij}, a_{ij'}$  from the corresponding row of  $\mathbf{A}$  such that  $a_{ij} - a_{ij'}$  is proportional to a modulated and translated Dirac comb with  $2k$  spikes.*

For filter matrices with a single row, since  $\tilde{\mathbf{A}} \neq \mathbf{A}$  means that the filters  $\tilde{a}_{1j}$  are permuted versions of  $a_{1j}$ , we obtain

**Corollary 2** *Consider  $\mathbf{A}$  with a single row ( $M = 1$ ). Assume that  $\tilde{\mathbf{A}} \neq \mathbf{A}$ , that (4) holds with*

$$2k \cdot \Delta(\tilde{\mathbf{A}}, \mathbf{A}) = L \quad (14)$$

*and that the filters in  $\mathbf{A}$  have disjoint supports (3). Then  $\|\tilde{\mathbf{A}}\|_0 > \|\mathbf{A}\|_0$ .*

### 4.3 Excessive pessimism?

The counter-examples built in Lemmata 3-4, which are associated to Dirac combs, are highly structured. They provide worst case well-posedness bounds, but existing probabilistic versions of uncertainty principles (see, e.g., the nice survey [15]) lead us to conjecture that if the sparse filters in  $\mathbf{A}$  are drawn at random (e.g. from Bernoulli-Gaussian distribution), the uniqueness guarantee of Theorem 2 will hold except with small probability  $O(L^{-\beta})$ , provided that  $k < c(\beta)L/\log L$ , for large  $L$ . This is left to further theoretical investigation.

## 5 Numerical experiments

The results achieved so far are theoretical well-posedness guarantee, but do not quite provide algorithms to compute the potentially unique (up to global permutation) solution of the frequency permutation problem. We conclude this paper with the description of a relatively naive optimization algorithm, an empirical assessment of its performance with Monte-Carlo simulations, and a discussion of how this compares with the theoretical uniqueness guarantees achieved above.

### 5.1 Proposed combinatorial algorithm

Given a “permuted” matrix  $\tilde{\mathbf{A}}$ , one wishes to find a set of frequency permutations yielding a new matrix  $\hat{\mathbf{A}}$  with minimum  $\ell^p$  norm.

The proposed algorithm starts from  $\hat{\mathbf{A}}_0 = \tilde{\mathbf{A}}$ . Given  $\hat{\mathbf{A}}_n$ , a candidate matrix  $\hat{\mathbf{A}}_{n+1,\pi}$  can be obtained by applying a permutation  $\pi$  at frequency  $\omega_n \equiv n \pmod{L}$ . Testing each possible permutation  $\pi$  and retaining the one  $\pi_n$  which minimises  $\|\hat{\mathbf{A}}_{n+1,\pi}\|_p$  yields the next iterate  $\hat{\mathbf{A}}_{n+1} := \hat{\mathbf{A}}_{n+1,\pi_n}$ . The procedure is repeated until the  $\ell^p$  norm  $\hat{\mathbf{A}}_n$  ceases to change. Since there is a finite number of permutations to try, the stopping criterion is met after sufficiently many iterations.

### 5.2 Choice of the $\ell^p$ criterion

In theory, it could happen that the stopping criterion is only met after a combinatorially large number of iterations. However, the algorithm stops much sooner in practice. In fact, if we were to use the  $\ell^0$  norm, the algorithm would typically stop after just one iteration, because the  $\ell^0$  norm attains its maximum value  $M \times N \times L$  for most frequency permutations except a few very special ones. For this reason, we chose to test the algorithm using  $\ell^p$  norms  $p > 0$ , which are not as “locally constant” as the  $\ell^0$  norm. To our surprise, the experiments below will show that the best performance is not achieved for small  $p$ , but rather for  $p = 2 - \epsilon$  with small  $\epsilon > 0$ . For  $p = 0$  and for  $p \geq 2$ , the algorithm indeed completely fails.

### 5.3 Monte-Carlo simulations

For various filter length  $L$ , sparsity levels  $k$  and dimensions  $M, N$ , random sparse filter matrices  $\mathbf{A}$  made of independent random  $k$ -sparse filters were generated. Each filter was drawn by choosing: a) a support of size  $k$  uniformly at random; b) i.i.d. Gaussian coefficients on this support.

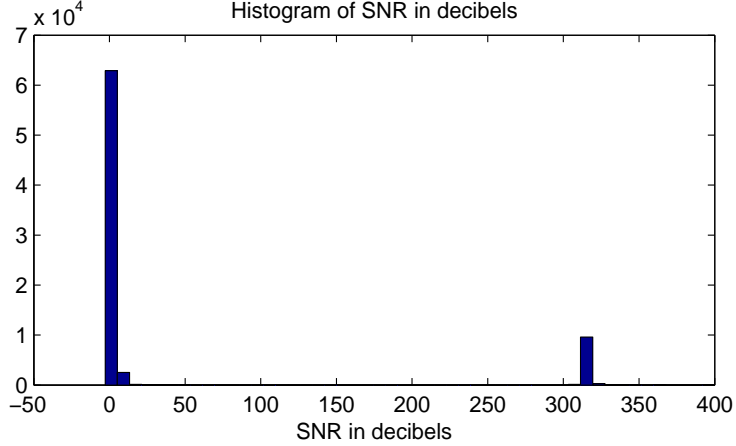


Figure 1: Histogram of SNR between best permutation of  $\hat{\mathbf{A}}$  and original  $\mathbf{A}$

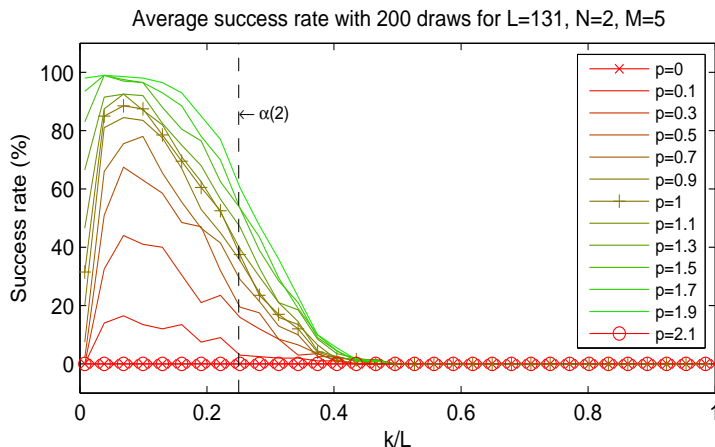
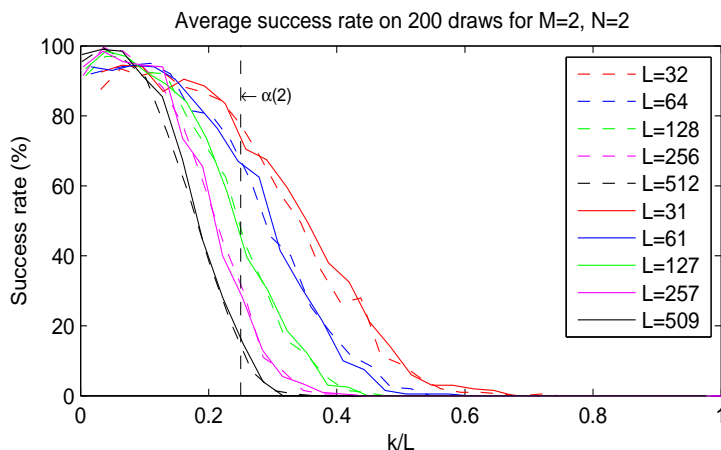
For each configuration  $(L, k, M, N)$ , 200 such random matrices  $\mathbf{A}$  were drawn. For each  $\mathbf{A}$ , independent random frequency permutations were applied to obtain  $\tilde{\mathbf{A}}$ . The algorithm was then applied to obtain  $\hat{\mathbf{A}}$ . The performance was measured using the SNR between  $\mathbf{A}$  and the best permutation of  $\hat{\mathbf{A}}$ .

Figure 1 shows the histogram of SNR values achieved for  $L = 31$ ,  $1 \leq k \leq L$ ,  $M \in \{1, 2\}$ ,  $N \in \{2, 3, 4\}$ ,  $p = 1$ . It shows that the algorithm either completely succeeds up to machine precision (SNR above 300 dB) or completely fails (SNR of the order of 0 dB). For this reason, in the rest of the experiments the estimation was considered a success when the SNR exceeded 100 dB.

#### 5.4 Role of the $\ell^p$ norm

Figure 2 displays the success rate as a function of the relative sparsity  $k/L$ , for various choices of the  $\ell^p$  criterion, with filters of prime length  $L = 131$ ,  $N = 2$  sources and  $M = 5$  channels. The vertical dashed line indicates the threshold  $k/L \leq \alpha(2)$  associated with the well-posedness guarantee (using an  $\ell^0$  criterion) of Theorem 2. Surprisingly, one can observe that the success rate increases when  $0 < p < 2$  is increased. The maximum success rate is achieved when  $p = 2 - \epsilon$  with small  $\epsilon > 0$ .

Beyond the well-posedness regime suggested by the theory (i.e., to the right of the vertical dashed line) the algorithm can succeed, but at a rate that rapidly decreases when the relative sparsity  $k/L$  increases. In the regime where the problem is proved to be well-posed, the proposed algorithm is often successful but can still fail to perfectly recover the filters, especially –and surprisingly– for small values of  $k$ . This phenomenon is strongly marked for  $p < 1$  and essentially disappears for  $p > 1$ . It remains an open question to determine the respective roles of the  $\ell^p$  criterion and of the naive greedy optimization algorithm in this limited performance when the problem is well-posed with respect to the  $\ell^0$  norm.

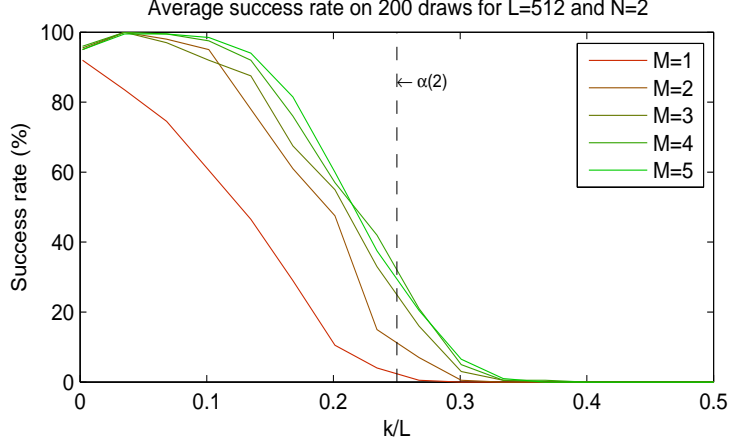
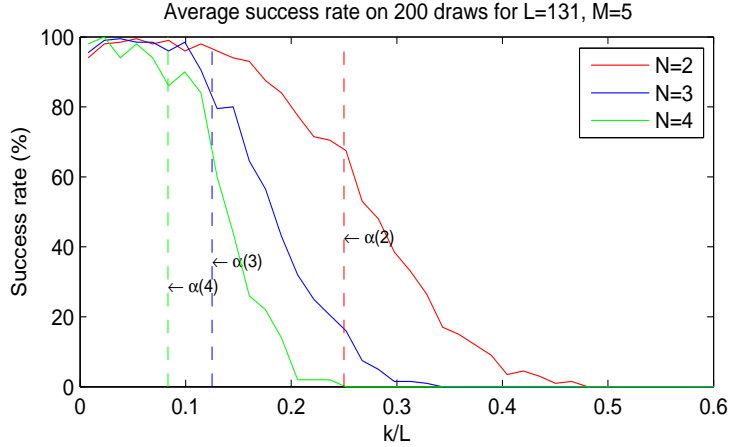

 Figure 2: Filter recovery success as a function of  $p$ ,  $0 \leq p \leq 1.9$ 

 Figure 3: Filter recovery success as a function of  $L$ , for  $p = 1.9$ 

## 5.5 Role of the filter length $L$

Figure 3 shows the results for different  $L$  values with  $p = 1.9$ ,  $M = N = 2$ . One can see that the average performance does not seem to depend on whether  $L$  is prime or not. As  $L$  increases, the performance for “small”  $k/L$  slightly increases, but the success rate degrades for “large”  $k/L$  close to  $\alpha(2)$ .

## 5.6 Role of the number of channels $M$

Figure 4 shows the results for increasing numbers of channels  $M$ , with a filter length  $L = 512$ ,  $N = 2$  sources,  $p = 1.9$ . One can observe that the success rate substantially increases when  $M$  is increased from  $M = 1$  to  $M = 2$ , and slightly increases as  $M$  further increases. Although the worst-case well-posedness guarantees are the same, the algorithm seems to benefit from added filter diversity across channels.

Figure 4: Filter recovery success as a function of  $M$ , for  $p = 1.9$ Figure 5: Filter recovery success as a function of  $N$ , for  $p = 1.9$ 

### 5.7 Role of the number of sources $N$

Figure 5 shows the success rate as a function of the relative sparsity  $k/L$ , for  $N \in \{2, 3, 4\}$ , with  $L = 31$ ,  $M = 5$  with  $p = 1.9$ . The well-posedness limits  $k/L \leq \alpha(N)$  associated to Theorem 2 are indicated with vertical dashed lines. The empirical curves confirm that the algorithm can still succeed beyond the worst-case well-posedness guarantees, but with a rapidly decreasing rate of success. When the well-posedness guarantees hold, the algorithm can fail, but its rate of success is high when the relative sparsity is sufficiently small compared to the bound provided by Theorem 2.

### 5.8 Computation time

The algorithm evaluates the  $\ell^p$  norm of the  $N!$  permutations of the sources for each of the  $L$  frequencies.

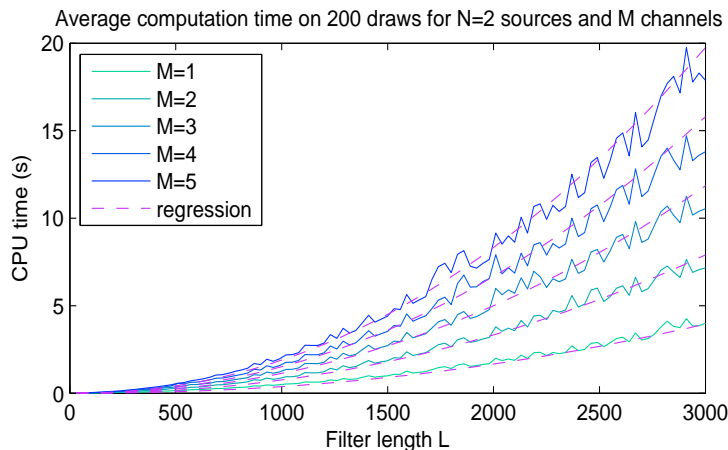


Figure 6: Computation time of the permutation solving algorithm depending on the length  $L$  of the filter

To evaluate the  $\ell^p$  norm of the filters, the permuted frequency coefficients have to be transformed back into the time domain by inverse Discrete Fourier transform. For each filter, the cost of the Discrete Fourier Transform through a Fast Fourier Transform is  $\mathcal{O}(L \log_2 L)$ . There are  $MN$  filters and hence the cost of  $\ell^p$  norm evaluation for a given configuration of sub-bands is  $\mathcal{O}(MNL \log_2 L)$ .

Hence, the complexity of each sweep through the set of all frequencies is  $\mathcal{O}(N!MNL^2 \log_2 L)$ . This is rather expensive because the computational cost grows in factorial with the number of sources and in square with the filter length, but it is tractable for small problem sizes and very efficient compared to the brute force approach that would require  $\mathcal{O}((N!)^{L-1}MNL \log_2 L)$  operations to test all  $(N!)^{L-1}$  possible permutations up to a global permutation.

Figure 6 shows the average computation time over 200 trials for various filter length. The red dashed line corresponds to its prediction using the theoretical cost estimation as  $C \times L^2 \log_2 L$  with  $C \approx 40$  nanoseconds.

## 6 Conclusions

It is now well known that a sufficient sparsity assumption can be used to make under-determined linear inverse problems well-posed: without the sparsity assumption, the problem admits an affine set of solutions, which intersects at only one point with the set of sparse vectors. Besides this well-posedness property, a key factor that has led to the large deployment of sparse models and methods in various fields of science is the fact that a convex relaxation of the NP-hard  $\ell^0$  minimization problem can be guaranteed to find this unique solution under certain sparsity assumptions. The availability of efficient convex solvers then really makes the problem tractable.

The problem considered in this paper is not a linear inverse problem. Even though it is a simplification of the original permutation *and scaling* problem arising from signal processing, it remains a priori a much harder problem than linear inverse problems in terms of the structure of the solution set: each solution



comes with a herd of solutions that are equivalent up to a global permutation.

The fact that we managed to obtain well-posedness results in this context is encouraging, but this is at best the beginning of the story: even if the solution is unique, how do we efficiently compute it? Can one hope to extend these results to the original permutation and scaling problem? Why does the proposed naive algorithm perform better for  $p > 1$ ? Answers to these questions are likely to have an impact in fields such as blind source separation with sparse multipath channels.

## A Proof of Theorem 1

First, notice that for each frequency  $\omega$  and channel  $i$ , permutations preserve the equality  $\sum_j a_{ij}[\omega] = \sum_j \tilde{a}_{ij}[\omega]$ . Thus, the same holds in the time-domain :  $\sum_j a_{ij} = \sum_j \tilde{a}_{ij}$ . By the disjoint supports hypothesis and the quasi-triangle inequality for  $\ell^p$  quasi-norms we have

$$\sum_j \|a_{ij}\|_p^p = \left\| \sum_j a_{ij} \right\|_p^p = \left\| \sum_j \tilde{a}_{ij} \right\|_p^p \leq \sum_j \|\tilde{a}_{ij}\|_p^p. \quad (15)$$

We conclude by summing over all channels  $i$ .

## B Proof of Lemma 2 and Corollary 1

Let us consider the matching count matrix  $\mathbf{C}$  with entries

$$C_{jn} := \#\{0 \leq \ell < L : \sigma_\ell(j) = n\}, \quad 1 \leq j, n \leq N.$$

Since  $\sum_j C_{jn} = \sum_n C_{jn} = L$  we have  $\mathbf{C} = L \cdot \mathbf{B}$  where  $\mathbf{B}$  is bi-stochastic.

A weakened version of Lemma 2, with  $2\alpha''(N) = \frac{1}{1+(N-1)^2}$ , can be obtained by combining the Birkhoff - Von Neumann theorem and Carathéodory theorem.

**Theorem 3 (Birkhoff - Von Neumann Theorem, [16, 17])** *Every bi-stochastic matrix is in the convex hull of permutation matrices.*

**Theorem 4 (Carathéodory Theorem [18])** *Let  $X$  be a non empty subset of an affine space of dimension  $n \geq 1$ . Then every element of the convex hull of  $X$  is a convex combination of  $p$  elements of  $X$ , with  $p \leq n + 1$ .*

The set of bi-stochastic matrices is an affine subspace of  $\mathbb{R}^{N^2}$ . It is defined by  $2N$  equations, but these equations are linearly dependent since the sum of the sum over rows is the sum of the sum over columns. Hence its affine dimension is  $n \leq N^2 - (2N - 1) = (N - 1)^2$ , and we conclude from Carathéodory's theorem that every bi-stochastic matrix is a convex combination of at most  $(N - 1)^2 + 1$  permutation matrices. One of the coefficients of this combination must therefore exceed  $1/(1 + (N - 1)^2)$ , and this leads to a version of Lemma 2 with  $2\alpha''(N) = \frac{1}{1+(N-1)^2}$ , as claimed.

Yet, this bound is suboptimal. The optimal bound in Lemma 2 follows from Hall's Marriage Theorem, which by the way is also a key ingredient in the proof of the Birkhoff-Von Neumann theorem.

**Theorem 5 (Hall's Marriage Theorem [14, 19] )** *Let  $(A_j)_{j \in J}$  be a family of subsets of a set finite  $S$ . There exists a bijection  $\pi : J \rightarrow S$  such that  $\pi(j) \in A_j$  for all  $J$  if, and only if, for all  $E \subset J$*

$$\# \cup_{j \in E} A_j \geq \#E$$

The bijection  $\pi$  is often referred to as a *transversal* for  $S$ .

**Proof (Lemma 2)** For shortness of notation we write  $\alpha$  for  $\alpha(N)$ . Define the sets  $J = S = \llbracket 1, N \rrbracket$ , and  $A_j := \{n : B_{jn} \geq 2\alpha\}$ ,  $j \in J$ , and consider the property

$$\mathcal{P}_k : \forall E \subset J, \#E \leq k \Rightarrow \# \cup_{j \in E} A_j \geq \#E.$$

We wish to prove that  $\mathcal{P}_k$  holds true for all  $1 \leq k \leq N$ : then, by Hall's Marriage Theorem, there exists a bijection  $\pi : j \rightarrow \pi(j)$  such that  $\pi(j) \in A_j$  for all  $j$ , yielding in turn the permutation matrix  $\mathbf{P}$  with ones at the entries  $(j, \pi(j))$ . We proceed by contradiction: assume that  $\mathcal{P}_N$  does not hold true. Since  $\mathcal{P}_1$  holds true, without loss of generality, for some  $1 \leq k_0 < N$ :

$$\# \cup_{1 \leq j \leq k_0} A_j \geq k_0, \text{ and } \# \cup_{1 \leq k \leq k_0+1} A_j \leq k_0.$$

Hence, without loss of generality:

$$\cup_{1 \leq k \leq k_0} A_j = \llbracket 1, k_0 \rrbracket \supset A_{k_0+1}.$$

It follows that for  $n > k_0$  and  $j \leq k_0 + 1$ , we have  $n \notin A_j$ , hence  $B_{jn} < 2\alpha$ . Now we use the bi-stochasticity of  $\mathbf{B}$  ( $\sum_j B_{jn} = \sum_n B_{jn} = 1$ ,  $B_{jn} \geq 0$ ) to obtain

$$\begin{aligned} k_0 &\geq \sum_{n \leq k_0} \sum_{j \leq k_0+1} B_{jn} = \sum_{j \leq k_0+1} \sum_{n \leq k_0} B_{jn} \\ &= \sum_{j \leq k_0+1} \left(1 - \sum_{n > k_0} B_{jn}\right) > \sum_{j \leq k_0+1} (1 - (N - k_0)2\alpha) \\ &= (k_0 + 1)(1 - (N - k_0)2\alpha) \\ &= k_0 + \left(1 - (k_0 + 1)(N - k_0)2\alpha\right). \end{aligned}$$

This implies  $2\alpha > 1/(k_0 + 1)(N - k_0)$ . However, this yields a contradiction, since a simple functional study shows that

$$\max_{1 \leq k_0 < N} \frac{1}{(k_0 + 1)(N - k_0)} = 2\alpha.$$

■

Equipped with Lemma 2, we can now prove Corollary 1.

**Proof (Corollary 1)** Since  $\mathbf{C} = L \cdot \mathbf{B}$  where  $\mathbf{B}$  is bi-stochastic, there is a permutation  $\pi$  such that  $C_{j\pi(j)} \geq 2L\alpha(N)$ .

■

We conclude this section by showing the sharpness of Corollary 1 through the construction of permutations that reach the bound. Consider  $N$  an integer,

and  $k_0 := N/2$  ( $N$  even) or  $k_0 := (N - 1)/2$  ( $N$  odd). Let  $L$  be a multiple of  $(k_0 + 1)(N - k_0)$ . Consider the  $L \times N$  matrix:

$$\Sigma := \left[ \begin{array}{ccccc|ccc} \mathbf{1} & \mathbf{U} & \mathbf{k}_0 & \dots & \mathbf{2} & \times & \dots & \times \\ \mathbf{2} & \mathbf{1} & \mathbf{U} & \ddots & \ddots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & & \vdots \\ \mathbf{k}_0 & \ddots & \ddots & \ddots & \mathbf{U} & \vdots & & \vdots \\ \mathbf{U} & \mathbf{k}_0 & \dots & \mathbf{2} & \mathbf{1} & \times & \dots & \times \end{array} \right]$$

where: a) the left part, of size  $L \times (k_0 + 1)$ , is filled with the column vectors  $\mathbf{i} \in \mathbb{R}^{L/(k_0+1)}$  made of constant entries equal to the integer  $1 \leq i \leq k_0$  and the vector  $\mathbf{U} \in \mathbb{R}^{L/(k_0+1)}$  made of the vertical concatenation of the  $N - k_0$  column vectors  $\mathbf{j} \in \mathbb{R}^{L/(k_0+1)(N-k_0)}$  with constant entries  $k_0 + 1 \leq j \leq N$ ; b) the rows of the right part, of size  $L \times (N - k_0 - 1)$ , include exactly once each integer  $1 \leq \ell \leq N$  which does not already appear in the corresponding row of the left part. By construction, the  $L$  rows of the matrix  $\Sigma$  are associated to  $L$  permutations  $\sigma_\ell$ . We now show that, for any global permutation  $\pi$ , there is at least one column  $1 \leq j \leq k_0 + 1$  such that

$$\#\{\ell : \sigma_\ell(j) = \pi(j)\} \leq L/(k_0 + 1)(N - k_0) = L2\alpha(N).$$

Applying again the pigeonhole principle yields: among the  $k_0 + 1$  indices  $j$  to consider, at least one,  $j^*$ , must be mapped to an integer  $\pi(j^*) \geq k_0 + 1$ . By construction, the columns of  $\Sigma$  are such that column  $j^*$  contains at most (in fact: exactly)  $L/(k_0 + 1)(N - k_0)$  instances of the value  $\pi(j^*)$ .

## C Proof of Theorem 2

We can now conclude the proof of Theorem 2. By Corollary 1, there is a permutation  $\pi$  such that for each  $j$ , we have  $\|\mathbf{F}(\tilde{a}_{ij} - a_{i\pi(j)})\|_0 \leq L(1 - 2\alpha(N))$ , hence  $\Delta(\tilde{\mathbf{A}}, \mathbf{A}|\pi) \leq L(1 - 2\alpha(N))$  and finally  $\Delta(\tilde{\mathbf{A}}, \mathbf{A}) \leq L(1 - 2\alpha(N))$ . Combined with the assumption  $k \leq L\alpha(N)$ , we obtain  $2k + \Delta \leq L$ , and we conclude thanks to Lemma 1.

## D Proof of Lemmata 1 3 4 5

We prove Lemma 1 first, then the statements of Lemma 3 in the following order: 1), 3), 2). We begin by some notations and fact regarding Dirac combs.

### D.1 Dirac combs

Let  $p, q \geq 1$  be two integers and  $L = pq$  their product. The unit Dirac comb with  $q$  spikes and of step  $p$ , denoted  $\mathfrak{w}_p$ , is the vector of  $\mathbb{C}^L$  defined by  $\mathfrak{w}_q[t] = 1/\sqrt{q}$  if  $t \equiv 0[p]$ ,  $\mathfrak{w}_q[t] = 0$  otherwise. Its Fourier transform is the unit Dirac comb with  $p$  spikes and of step  $q$ :  $\mathbf{F}\mathfrak{w}_q = \mathfrak{w}_p$ . For  $0 \leq n < p$  an integer translation index and  $0 \leq m < q$  an integer modulation index, one can define the translated and modulated Dirac comb  $\mathfrak{w}_{q,n,m} = T_n M_m \mathfrak{w}_q$  where  $T_n$  is the circular shift by  $n$  samples, and  $M_m$  is the frequency modulation  $(M_m u)[t] := u[t] \cdot e^{2i\pi mt/L}$ .

One can check that the collection  $\{\mathfrak{m}_{q,n,m}\}_{0 \leq n < p, 0 \leq m < q}$  is an orthonormal basis of  $\mathbb{C}^L$ .

## D.2 Proof of Lemma 1

Let  $\pi_0$  be the permutation such that  $\Delta_0(\mathbf{A}, \tilde{\mathbf{A}}) = \min_{\pi \in \mathfrak{S}_N} \Delta_0(\mathbf{A}, \tilde{\mathbf{A}}|\pi)$ . By abuse of notation we still denote  $\mathbf{A}$  the matrix obtained by applying  $\pi_0$  to permute the columns of the original filter matrix. For each channel  $i$  and a source index  $j$  such that  $a_{ij} = \tilde{a}_{ij}$  we obviously have  $\|a_{ij}\|_0 \leq \|\tilde{a}_{ij}\|_0$ . Now, since  $\Delta_0 \geq 1$  we have  $\tilde{\mathbf{A}} \neq \mathbf{A}$  hence there exists a pair  $i, j$  such that  $\tilde{a}_{ij} \neq a_{ij}$ . By the  $\ell^0$  Dirac-Fourier uncertainty principle [12, Theorem 1], for any vector  $u \in \mathbb{C}^L$  we have  $\|u\|_0 \|\mathbf{F}u\|_0 \geq L$ . Hence, by the hypothesis  $k < L/(2\Delta_0)$  we have

$$\|a_{ij}\|_0 + \|\tilde{a}_{ij}\|_0 \geq \|\tilde{a}_{ij} - a_{ij}\|_0 \quad (16)$$

$$\geq L/\|\mathbf{F}(\tilde{a}_{ij} - a_{ij})\|_0 \quad (17)$$

$$\geq L/\Delta_0 > 2k \quad (18)$$

$$\geq \|a_{ij}\|_0 + \|a_{ij'}\|_0 \quad (19)$$

where  $j'$  is an arbitrary source index. Hence for every  $i, j$  such that  $\tilde{a}_{ij} \neq a_{ij}$  and any  $j'$ ,  $\|\tilde{a}_{ij}\|_0 > \|a_{ij'}\|_0$ , and we obtain

$$\|\tilde{a}_{ij}\|_0 > \max_{j'} \|a_{ij'}\|_0 \geq \|a_{ij}\|_0.$$

Overall, we have shown that  $\|\tilde{\mathbf{A}}\|_0 > \|\mathbf{A}\|_0$ .

When  $L$  is prime, a stronger uncertainty principle  $\|u\|_0 + \|\mathbf{F}u\|_0 \geq L + 1$  holds [13]. Hence, under the assumption  $2k + \Delta_0 \leq L$  we can replace (17)-(18) with

$$\dots \geq L + 1 - \|\mathbf{F}(\tilde{a}_{ij} - a_{ij})\|_0 \geq L + 1 - \Delta_0 > 2k$$

to reach the same conclusion.

## D.3 Proof of Lemma 3

We shall simply build an example where  $\mathbf{A} = [\alpha, \beta]$  is a  $1 \times 2$  matrix of filters. Extensions to  $\mathbf{A}$  an  $M \times N$  matrix are trivial by adding mutually distinct sparse columns that are distinct from  $\alpha$  and  $\beta$ , and duplicating the first row.

We exploit Dirac combs as described in Appendix D.1. Define  $a = \mathfrak{m}_{k,0,0}$ ,  $b = -\mathfrak{m}_{k,L/2k,0}$ . The filters  $a$  and  $b$  have disjoint support and satisfy  $\|a\|_0 = \|b\|_0 = k$ . Since  $a - b = \sqrt{2} \mathfrak{m}_{2k,0,0}$  we have  $a[\omega] = b[\omega]$  whenever  $\omega \not\equiv 0[2k]$ . Hence, permuting the Fourier transforms of  $a$  and  $b$  on the  $L/2k$  frequencies  $\{\omega = 2kr, 0 \leq r < L/2k\}$  yields  $\tilde{a} = b$  and  $\tilde{b} = a$ . Given any  $u \in \mathbb{C}^L$  we define perturbations  $\alpha$  and  $\beta$  of  $a$  and  $b$

$$\begin{cases} \alpha & := a + u \\ \beta & := b + T_{L/2}u \end{cases}$$

with  $T_{L/2}$  a circular shift. Noticing that for  $\omega = 2kr$

$$(T_{L/2}u)[\omega] = e^{\frac{2i\pi(L/2)\omega}{L}} u[\omega] = e^{2i\pi kr} u[\omega] = u[\omega]$$

we obtain that, after permuting the Fourier transforms of  $\alpha$  and  $\beta$  at the frequencies  $\omega = 2kr$ ,  $0 \leq r < L/2k$ ,

$$\begin{cases} \tilde{\alpha} &= b + u \\ \tilde{\beta} &= a + T_{L/2}u \end{cases}$$

We choose the vector  $u$  to be zero everywhere with two exceptions  $u[0] := -a[0]$ ,  $u[\frac{L}{2k}] := -b[\frac{L}{2k}]$ . Since  $T_{L/2}u \neq u$  and  $a \neq b$ , we have  $\{\alpha, \beta\} \neq \{\tilde{\alpha}, \tilde{\beta}\}$  and  $\tilde{\mathbf{A}} \neq \mathbf{A}$ . Moreover,  $\Delta_0(\tilde{\mathbf{A}}, \mathbf{A}) = \Delta_1(\tilde{\mathbf{A}}, \mathbf{A}) = L/2k$ .

Lastly, all considered vectors have  $k$  entries of equal magnitude, hence  $\|\alpha\|_0 = \|\beta\|_0 = \|\tilde{\alpha}\|_0 = \|\tilde{\beta}\|_0 = k$ , and for any  $0 < p \leq \infty$   $\|\alpha\|_p = \|\beta\|_p = \|\tilde{\alpha}\|_p = \|\tilde{\beta}\|_p$ . In particular,  $\|\tilde{\mathbf{A}}\|_p = \|\mathbf{A}\|_p$ ,  $0 \leq p \leq \infty$ .

#### D.4 Proof of Lemma 5

We repeat the construction of the proof of Lemma 3 starting from the Dirac combs  $a = \mathfrak{m}_{k',0,0}$ ,  $b = -\mathfrak{m}_{k',L/2k',0}$ . Since  $k' < k \leq L/2$ , we have  $\ell := k - k' \leq L/2 - k'$  hence we can choose an  $\ell$ -sparse vector  $u$  which support is outside the support of  $\mathfrak{m}_{2k'}$  and such that  $T_{L/2}u$  and  $u$  have disjoint supports. The four vectors  $\{a, b, u, T_{L/2}u\}$  have mutually disjoint supports, hence  $\alpha$  and  $\beta$  have disjoint supports,  $\{\alpha, \beta\} \neq \{\tilde{\alpha}, \tilde{\beta}\}$  and  $\tilde{\mathbf{A}} \neq \mathbf{A}$ . Moreover,  $\Delta_0(\tilde{\mathbf{A}}, \mathbf{A}) = \Delta_1(\tilde{\mathbf{A}}, \mathbf{A}) = L/2k$ . Lastly, we have  $\|\alpha\|_0 = \|\beta\|_0 = \|\tilde{\alpha}\|_0 = \|\tilde{\beta}\|_0 = k' + \ell = k$ , and the  $\ell^p$  norms of these vectors are also equal, hence  $\|\tilde{\mathbf{A}}\|_p = \|\mathbf{A}\|_p$ ,  $0 \leq p \leq \infty$ .

#### D.5 Proof of Lemma 4

As in the proof of Lemma 1 we consider  $\mathbf{A}$  the permuted matrix associated to the optimal permutation  $\pi_0$ . Using the inequality  $2k \leq L/\Delta_1 \leq L/2\Delta_0$  instead of  $2k < L/\Delta_0$  we repeat the steps (16)-(19) to obtain  $\|\tilde{a}_{ij}\|_0 \geq \|a_{ij'}\|_0$  for any  $j \in E_i := \{j, a_{ij} \neq \tilde{a}_{ij}\}$  and any  $j'$ . As a result  $\|\tilde{a}_{ij}\|_0 \geq \|a_{ij}\|_0$  for all  $i, j$ . The assumption that  $\|\tilde{\mathbf{A}}\|_0 = \|\mathbf{A}\|_0$  implies that  $\|\tilde{a}_{ij}\|_0 = \|a_{ij}\|_0$  for all  $i, j$ .

By assumption,  $\tilde{\mathbf{A}} \neq \mathbf{A}$  hence there are indices  $i, j$  such that  $a_{ij} \neq \tilde{a}_{ij}$ . For such  $i, j$ , since  $\|a_{ij}\|_0 = \|\tilde{a}_{ij}\|_0$ , each inequality in (16)-(19) (the inequality  $L/\Delta_0 > 2k$  being replaced with  $L/\Delta_1 \geq 2k$ ) must be indeed an equality. This implies that:  $\|a_{ij}\|_0 = \|\tilde{a}_{ij}\|_0 = k$ ;  $2k$  divides  $L$  and  $\Delta_1 = L/2k$ ; the nonzero vector  $b_{ij} := \tilde{a}_{ij} - a_{ij}$  must be an equality case of the  $\ell^0$  uncertainty principle with  $\|b_{ij}\|_0 = 2k$  and  $\|\mathbf{F}b_{ij}\|_0 = L/2k$ . As a result [13]  $b_{ij}$  is a scaled, modulated and translated version of the Dirac comb  $\mathfrak{m}_{2k}$  made of  $2k$  Diracs spaced every  $L/2k$  samples: there exists a scalar  $\gamma_{ij} \neq 0$ , and two integers  $0 \leq n_{ij} < L/2k$ ,  $0 \leq m_{ij} < 2k$  such that

$$b_{ij} = \gamma_{ij} \cdot \mathfrak{m}_{2k, n_{ij}, m_{ij}}.$$

Moreover since  $\|a_{ij}\|_0 = \|\tilde{a}_{ij}\|_0 = k$  and  $\|\tilde{a}_{ij} - a_{ij}\|_0 = 2k$ , the filters  $\tilde{a}_{ij}$  and  $a_{ij}$  have disjoint supports of size  $k$ . Hence, they are the restriction of  $b_{ij}$  (resp. of  $-b_{ij}$ ) to their respective supports.

Now, define

$$E_{i,n,m} := \{j \in E_i, n_{ij} = n, m_{ij} = m\}.$$

As observed in the proof of Theorem 1, the equality  $\sum_j a_{ij} = \sum_j \tilde{a}_{ij}$  holds, implying  $\sum_{j \in E_i} b_{ij} = \sum_j b_{ij} = 0$ . Taking inner products with the Dirac comb orthonormal basis  $\mathfrak{m}_{2k,n,m}$ ,  $0 \leq n < L/2k$ ,  $0 \leq m < 2k$ , yields

$$\sum_{j \in E_{i,n,m}} \gamma_{ij} = 0, \quad (20)$$

Since  $\gamma_{ij} \neq 0$ , whenever  $E_{i,n,m}$  is not empty it contains at least two distinct indices.

By the *disjoint support assumption*: for  $j, j' \in E_{i,n,m}$ ,  $j \neq j'$ , the original filters  $a_{ij}$  and  $a_{ij'}$  have disjoint supports. Moreover, we know that these supports are subsets of the support of  $\mathfrak{m}_{2k,n,m}$  which is of size  $2k$ , hence

$$\#E_{i,n,m} \cdot k = \left\| \sum_{j \in E_{i,n,m}} a_{ij} \right\|_0 \leq 2k.$$

Hence, whenever  $E_{i,n,m}$  is not empty, it contains *exactly* two distinct elements:  $E_{i,n,m} = \{j, j'\}$  where  $j \neq j'$ .

Further, observe that: a)  $a_{ij}$  and  $a_{ij'}$  have disjoint supports of size  $k$  which are subsets of the support of size  $2k$  of  $\mathfrak{m}_{2k,n,m}$ ; b)  $a_{ij}$  and  $\tilde{a}_{ij}$  have the same property. As a result,  $\tilde{a}_{ij}$  and  $a_{ij'}$  have the same support, which is disjoint from that of  $a_{ij}$ . Similarly,  $a_{ij}$  has the same support as  $\tilde{a}_{ij'}$ . Finally, Eq. (20) can be rewritten  $\gamma_{ij} + \gamma_{ij'} = 0$ , and implies  $b_{ij} + b_{ij'} = 0$ , that is to say  $\tilde{a}_{ij} + \tilde{a}_{ij'} = a_{ij'} + a_{ij}$ . We conclude that  $\tilde{a}_{ij} = a_{ij'}$  and  $\tilde{a}_{ij'} = a_{ij}$ .

## Acknowledgment

This work was supported by the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL: Sparse Models, Algorithms and Learning for Large-Scale data, and by Agence Nationale de la Recherche (ANR), project ECHANGE (ANR-08-EMER-006).

## References

- [1] C.R. Berger, Shengli Zhou, J.C. Preisig, and P. Willett. Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing. *Signal Processing, IEEE Transactions on*, 58(3):1708–1721, March 2010.
- [2] M. Sharp and A. Scaglione. Application of sparse signal recovery to pilot-assisted channel estimation. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 3469–3472, April 2008.
- [3] W.U. Bajwa, J. Haupt, A.M. Sayeed, and R. Nowak. Compressed channel sensing: A new approach to estimating sparse multipath channels. *Proceedings of the IEEE*, 98(6):1058–1076, June 2010.
- [4] Jacob Benesty, M. Mohan Sondhi, and Yiteng (Arden) Huang. *Springer Handbook of Speech Processing*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

- [5] Binning Chen and A.P. Petropulu. Frequency domain blind mimo system identification based on second and higher order statistics. *Signal Processing, IEEE Transactions on*, 49(8):1677–1688, Aug 2001.
- [6] Pierre Comon and Christian Jutten, editors. *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. Academic Press, 2010.
- [7] M.S. Pedersen, J. Larsen, U. Kjems, and L.C. Parra. A survey of convolutive blind source separation methods. *Multichannel Speech Processing Handbook*, 2007.
- [8] Christine Serviere and Dinh-Tuan Pham. A novel method for permutation correction in frequency-domain in blind separation of speech mixtures. In Carlos Puntonet and Alberto Prieto, editors, *Independent Component Analysis and Blind Signal Separation*, volume 3195 of *Lecture Notes in Computer Science*, pages 807–815. Springer Berlin / Heidelberg, 2004.
- [9] S. Sanei, Wenwu Wang, and J.A. Chambers. A coupled hmm for solving the permutation problem in frequency domain bss. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, volume 5, pages V – 565–8 vol.5, may 2004.
- [10] Wenwu Wang, Jonathon A. Chambers, and Saeid Sanei. A novel hybrid approach to the permutation problem of frequency domain blind source separation. In *ICA '04*, pages 532–539, 2004.
- [11] David L. Donoho and P.B. Stark. Uncertainty principles and signal recovery. *SIAM Journal on Applied Mathematics*, 49(3):906–931, 1989.
- [12] M. Elad and A.M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *Information Theory, IEEE Transactions on*, 48(9):2558–2567, 2002.
- [13] Terence Tao. An uncertainty principle for cyclic groups of prime order. *Mathematical Research Letters*, 12:121–127, 2005.
- [14] Philip Hall. On representatives of subsets. *J. London Math. Soc.*, 10(1):26–30, 1935.
- [15] Joel Tropp. On the linear independence of spikes and sines. *Journal of Fourier Analysis and Applications*, 14:838–858, 2008. 10.1007/s00041-008-9042-0.
- [16] G. Birkhoff. Three observations on linear algebra, univ. *Nac. Tucumán. Rev. Ser. A*, 5:147–151, 1946.
- [17] J. Von Neumann. A certain zero-sum two-person game equivalent to the optimal assignment problem. *Contributions to the Theory of Games*, 2:5–12, 1953.
- [18] M. Berger. *Géométrie*, volume 4. Cedic, 1977.
- [19] J. Oxley. *Matroid theory*, volume 21. Oxford University Press, USA, 1992.



**RESEARCH CENTRE  
RENNES – BRETAGNE ATLANTIQUE**

Campus universitaire de Beaulieu  
35042 Rennes Cedex

Publisher  
Inria  
Domaine de Voluceau - Rocquencourt  
BP 105 - 78153 Le Chesnay Cedex  
[inria.fr](http://inria.fr)

ISSN 0249-6399